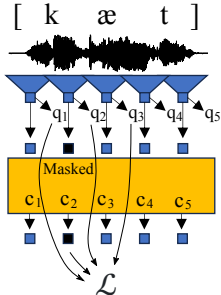


Feature extractor (49 Hz)

Quantiser

Transformer



Contrastive Loss + Diversity Loss